

A robust optimization approach to experimental design for model discrimination of dynamical systems

Dominik Skanda¹ Dirk Lebiedz²

February 2, 2012

Abstract

A high-ranking goal of interdisciplinary modeling approaches in science and engineering are quantitative prediction of system dynamics and model based optimization. Quantitative modeling has to be closely related to experimental investigations if the model is supposed to be used for mechanistic analysis and model predictions. Typically, before an appropriate model of an experimental system is found different hypothetical models might be reasonable and consistent with previous knowledge and available data. The parameters of the models up to an estimated confidence region are generally not known *a priori*. Therefore one has to incorporate possible parameter configurations of different models into a model discrimination algorithm which leads to the need for robustification. In this article we present a numerical algorithm which calculates a design of experiments allowing optimal discrimination of different hypothetical candidate models of a given dynamical system for the most inappropriate (worst case) parameter configurations within a parameter range. The design comprises initial values, system perturbations and the optimal placement of measurement time points, the number of measurements as well as the time points are subject to design. The statistical discrimination criterion is worked out rigorously for these settings, a derivation from the Kullback-Leibler divergence as optimization objective is presented for the case of discontinuous Heaviside-functions modeling the measurement decision which are replaced by continuous approximation during the optimization procedure. The resulting problem can be classified as a semi-infinite optimization problem which we solve in an outer approximations

¹University of Freiburg
 Center for Systems Biology (ZBSA)
 Habsburgerstr. 49
 79104 Freiburg
 Germany

² University of Freiburg
 Center for Systems Biology (ZBSA)
 Faculty of Biology
 and Department of Applied Mathematics,
 Habsburgerstr. 49
 79104 Freiburg
 Germany

approach stabilized by a suggested homotopy strategy whose efficiency is demonstrated. We present the theoretical framework, algorithmic realization and numerical results.

1 Introduction

High-ranking goals of interdisciplinary modeling approaches in the natural sciences are quantitative prediction of system dynamics and model based optimization. In particular in modern systems biology a related issue is to link molecular attributes to dynamic mechanisms and functional properties at the system level in order to mechanistically understand emerging functionality. For these purposes, mathematical modeling, numerical simulation and scientific computing techniques are indispensable. Quantitative modeling closely combined with experimental investigations is required if the model is supposed to be used for sound mechanistic analysis and model predictions.

Typically, before an appropriate model of a system is found different hypothetical models might be reasonable and consistent with previous knowledge and available data. The goal of this article is to derive, develop, implement and apply a numerical algorithm which calculates in a suitable sense an optimal design of experiments which allows the best discrimination of different hypothetic candidate models in form of ordinary differential equations (ODE). The algorithmic idea is to iteratively separate the response of different models by use of variations of experimental conditions and perturbations to the system.

To discriminate a set of candidate models against a given set of experimental data likelihood ratio tests based on bootstrap methods have been described in the literature, see e.g. [22], [46] or [48]. Ranking methods like Stewart's method ([44]) or the well known Akaike information criterion (see e.g. [15]) are popular as well in the field of biological modeling. Applications can be found for example in [24] or [10].

In contrast to these approaches our work deals with the problem of designing experiments so that statistical methods can be exploited in an optimal sense for model discrimination. This differs from the approach to find an experimental design to best estimate the parameters of a model for a given experimental system in terms of criteria characterizing the confidence regions [25, 7, 6].

Different approaches to design experiments for model discrimination exist. Besides optimization methods (see e.g [29], [19], [47] or [26]) a model-based feedback controller see e.g [3] and Markov chain Monte Carlo sampling methods [34] have been used to construct an appropriate design. An overview of various experimental design techniques can be found in [27].

Here, we present a robust numerical optimization algorithm which calculates the optimal design of experiments allowing the best discrimination of different candidate ODE models. An appropriate model and its parameters up to an estimated confidence region are not known *a priori*. Therefore one has to incorporate possible parameter configurations of different models into a model discrimination algorithm. The aim is to calculate the most discriminable response of different models for the most inappropriate parameter configurations within a parameter range via a worst case estimate. In that context inappropriate parameter configurations refer to the case when different models have calibrated parameter values such that the models have the most similar response. This worst case

estimate leads to the formulation of a maxmin optimization problem. Building on our previous work [43] we present an algorithm to compute robust optimal experimental designs. For the robustification we set up an outer approximation approach stabilized by a homotopy strategy.

The article is organized as follows. In Section 2.1 we give a brief overview of so called Kullback-Leibler(KL)-optimality as discussed by López-Fidalgo et al. [32] in the context of model discrimination. In Section 2.2 we derive our optimal design criterion by use of KL-optimality. In Section 3 the theoretical framework for the calculation of a robust design via solution of a maxmin optimization problem is presented. The numerical implementation is discussed in Section 3.1. The homotopy solution strategy is presented in Section 3.2. We demonstrate applications of the algorithmic framework for two test cases from biology, an allosteric metabolic enzyme model for glycolytic oscillations and a model describing signal sensing in *dictyostelium discoideum*, results are presented in Section 4.

2 Statistical Basis of Model Discrimination

2.1 KL-optimal design

In this section a model discrimination criterion based on the Kullback-Leibler (KL) divergence called KL-optimality as discussed by López-Fidalgo et al. [32] is introduced. López-Fidalgo et al. [32] demonstrate that KL-optimality is consistent with T-optimality [5] and generalized T-optimality [49] which are well known model discrimination criteria based on statistical tests.

We introduce the concept of a probability space and formally define the KL-divergence.

Definition 1. *The probability space is a triple (Ω, \mathcal{F}, P) consisting of*

- *a non-empty set Ω (sample space),*
- *a σ -algebra $\mathcal{F} \subseteq \mathcal{P}(\Omega)$, $E \in \mathcal{F}$ is called an event*
- *a probability measure $P : \mathcal{F} \rightarrow [0, 1]$.*

Definition 2. *Two probability spaces $(\Omega, \mathcal{F}, P_i)$, $i = 1, 2$, are called absolutely continuous with respect to each other, in symbols $P_1 \equiv P_2$, if $\nexists E \in \mathcal{F} : (P_1(E) = 0 \text{ AND } P_2(E) \neq 0) \text{ OR } (P_1(E) \neq 0 \text{ AND } P_2(E) = 0)$.*

The Radon-Nikodym Theorem allows a representation of a probability measure via a measurable probability density function.

Theorem 1. *(Radon-Nikodym)*

Let λ be a probability measure such that $\lambda \equiv P_1$, $\lambda \equiv P_2$. Then λ -measurable functions $f_i : \Omega \rightarrow \mathbb{R}$, $i = 1, 2$, called generalized probability densities, exist which are unique up to sets of measure zero and non-negative, such that

$$P_i(E) = \int_E f_i(x) d\lambda(x), \quad i = 1, 2, \quad (1)$$

for all $E \in \mathcal{F}$.

A proof of this theorem can be found e.g in [12].

In the following we use X for the generic variable and x for a specific value of X . If H_i , $i = 1, 2$ is the hypothesis that X is from the statistical population with probability measure P_i , the mean information for discrimination in favor of H_1 against H_2 given $x \in E \in \mathcal{F}$, for P_1 is given by the Kullback–Leibler divergence.

Definition 3. *The Kullback–Leibler (KL) divergence is given by*

$$\begin{aligned} \mathcal{I}(P_1 : P_2; E) &:= \frac{1}{P_1(E)} \int_E \log \frac{f_1(x)}{f_2(x)} dP_1(x) \\ &= \begin{cases} \frac{1}{P_1(E)} \int_E f_1(x) \log \frac{f_1(x)}{f_2(x)} d\lambda(x), & \text{if } P_1(E) > 0, \\ 0, & \text{if } P_1(E) = 0, \end{cases} \end{aligned} \quad (2)$$

with

$$dP_1(x) = f_1(x) d\lambda(x). \quad (3)$$

When E is the entire sample space Ω , we shorten the notation to $\mathcal{I}(P_1 : P_2)$. For discrete sets E the integral is substituted by a sum.

For details we refer to [28].

Now assume that the sample space Ω is split into two disjoint sets E_1 and E_2 , $\Omega = E_1 \cup E_2$. We define a statistical test procedure to choose between hypotheses H_1 and H_2 by accepting H_1 if $x \in E_1$ and accepting H_2 if $x \in E_2$. Assuming that one of the hypotheses has to be true we treat H_2 as the null hypothesis and call E_1 the critical region. The following wrong test decisions can occur.

Definition 4. *Incorrectly accepting H_1 although H_2 is true is called the type I error. The probability that this error occurs is given by*

$$\alpha = \text{Prob}(x \in E_1 | H_2) = P_2(E_1). \quad (4)$$

Definition 5. *Incorrectly accepting H_2 although H_1 is true is called the type II error. The probability that this error occurs is given by*

$$\beta = \text{Prob}(x \in E_2 | H_1) = P_1(E_2). \quad (5)$$

We assume that the test is repeated n -times and denote by \mathcal{O}_n a sample of n independent observations. \mathcal{O}_1 represents a sample of a single observation. β_n is defined as the corresponding probability of an error of type II which depends on the number of independent observations and the splitting of the probability space Ω into disjoint sets E_1 and E_2 .

The following theorem demonstrates an asymptotic relation between the KL-divergence and the minimum possible probability β_n^* of an error of type II with respect to all possible splittings $E_1 \cup E_2 = \Omega$ with given $\alpha = \text{Prob}(x \in E_1 | H_2) = P_2(E_1)$ [18].

Theorem 2. For any value of α , say α_0 , $0 < \alpha_0 < 1$,

$$\lim_{n \rightarrow \infty} (\beta_n^*)^{1/n} = e^{-\mathcal{I}(P_2 : P_1, \mathcal{O}_1)} \quad (6)$$

A proof of this theorem is given in [18, 28].

Assuming probability models for the outcome of a data measurement experiment depending on experimental design parameters $\xi \in \Xi \subset \mathbb{R}^d$, this theorem justifies the KL-divergence to be an appropriate objective functional for model-based computation of an optimal experimental design for discrimination between model hypotheses. For a design with the largest possible value of \mathcal{I} the asymptotic probability β_n^* of encountering an error of type II becomes minimal with respect to all possible splittings $E_1 \cup E_2 = \Omega$ with given α_0 . We indicate the dependency of the KL-divergence on the design by $\mathcal{I}(P_2 : P_1, \mathcal{O}_1; \xi)$. Our aim is to derive an algorithm to calculate the optimal design $\hat{\xi} \in \Xi$ such that

$$\hat{\xi} = \arg \max_{\xi \in \Xi} \mathcal{I}(P_2 : P_1, \mathcal{O}_1; \xi). \quad (7)$$

An extension of the case to test a simple null hypothesis against a simple alternative hypothesis to the more general case of both hypotheses being composite is generally of interest. This includes the situation to test whether given measurement data can be explained best by one out of a finite set of probability models based on measures P_{r_1} , $r_1 \in \{1, \dots, M_1\}$ parametrized by parameters $\theta_{r_1} \in \Theta_{r_1}$ where $\Theta_{r_1} \subset \mathbb{R}^{p_{r_1}}$ is the set of all possible parameter values to parametrize P_{r_1} and M_1 is the cardinality of the set of probability models, against the hypothesis that the measurement can best be explained by another one out of a second finite set of probability models based on measures P_{r_2} , parametrized by parameters $\theta_{r_2} \in \Theta_{r_2}$ where $\Theta_{r_2} \subset \mathbb{R}^{p_{r_2}}$ and $r_2 \in \{1, \dots, M_2\}$.

By calculating

$$\hat{\xi} = \arg \max_{\xi \in \Xi} \min_{\substack{r_1 \in \{1, \dots, M_1\} \\ r_2 \in \{1, \dots, M_2\}}} \min_{\substack{\theta_{r_1} \in \Theta_{r_1} \\ \theta_{r_2} \in \Theta_{r_2}}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi) \quad (8)$$

we can get a robust worst case estimate of an optimally discriminating design for the case of composite null and alternative hypothesis.

In practical applications a simple strategy to sort different probability models into null and alternative hypothesis would be to first rank all models according to the existing measurements and then set the best model as null hypothesis and the others as alternative hypothesis. The development of a suitable and efficient strategy is subject to further work.

It should be noted that the presented criteria is not symmetric, i.e. the null hypothesis is favored. In the case that both hypotheses might be equally reasonable we suggest to use the symmetrized version of the KL-divergence, i.e.

$$\mathcal{I}_{\text{sym}}(P_2 : P_1, \mathcal{O}_1; \xi) = \frac{\mathcal{I}(P_2 : P_1, \mathcal{O}_1; \xi) + \mathcal{I}(P_1 : P_2, \mathcal{O}_1; \xi)}{2}, \quad (9)$$

as optimization objective instead.

2.2 Derivation of the optimal experimental design criterion

In this section we derive a numerically computable optimization objective functional based on the framework of KL divergence. The derivation is motivated by the requirements of biological in vitro time series experiments modeled by kinetic ODE systems. In most situations such experiments are time and cost consuming. Therefore a central issue is to get the most information out of a single time series data measurement experiment taking place within a given fixed time span $[0, T_{\text{end}}]$. This means that in an optimal experimental design the most informative measurement time points for one measurement run have to be calculated in such a way that only one measurement at one time point can be performed. Often, an experiment cannot produce measurements in a time continuous way. Therefore we assume that there has to be a minimal time span ΔT for the separation of subsequent measurement time points. This contrasts to the usual approach to associate weights to a discrete or continuous time design scheme, see e.g. [5]. Additionally, the initial species concentrations of the participating species should be chosen in a most discriminating way.

A commonly used experimental practice is to combine kinetic time series measurements with perturbation stimuli like external adding of species quantities. From the model discrimination point of view the optimal time point of perturbations and the optimal species quantities to be added should be determined. We further assume that a measurement cannot be done at the same time as a perturbation.

In the following we translate these experimental conditions into a statistical model. Given the measurement time-vector $t \in \mathbb{R}_+^n$ with entries t_i for the n measurement time points $t_i, i \in \{1, \dots, n\}$ such that $t_{i+1} \geq t_i$, the “internal” model response vectors $y_{r_j}^i$ at measurement time t_i for the parametrized probability measures P_{r_j} , $r_j \in \{1, \dots, M_j\}$, $j \in \{1, 2\}$ are given by

$$y_{r_j}^i := y_{r_j}(t_{i-1}, t_i, y_{r_j}^{i-1} + c_{i-1}, \theta_{r_j}), \quad (10)$$

where $y_{r_j}(t_{i-1}, t_i, y_{r_j}^{i-1} + c_{i-1}, \theta_{r_j}) \in \mathbb{R}^{m_{r_j}}$ are the solutions of the initial value problems

$$\frac{dy_{r_j}}{dt} = f_{r_j}^{\text{rhs}}(y_{r_j}, \theta_{r_j}), \quad t \in [t_{i-1}, t_i], \quad (11)$$

with initial state $y_{r_j}(t^{i-1}) := y_{r_j}^{i-1} + c_{i-1}$ at end time t_i where $t_0 := 0$ and $c_0 := 0$. The vectors $c_i \in \mathbb{R}^{m_{\max}}$ denote species quantities the system can be perturbed with at time points t_i where m_{\max} is the maximum dimension of the models, i.e.

$$m_{\max} := \max_{j \in \{1, 2\}} \max_{r \in \{1, \dots, M_j\}} m_{r_j}. \quad (12)$$

$f_{r_j}^{\text{rhs}}(\cdot, \cdot)$ are the right hand side functions of the ODE models. $y_I \in \mathbb{R}^{m_{\max}}$ denotes the initial species concentration vector of the entire experiment which is for all models the same, i.e. $y_{r_j}^0 := y_I$. By m_{\min} we denote the minimal dimension of the models, i.e.

$$m_{\min} := \min_{j \in \{1, 2\}} \min_{r \in \{1, \dots, M_j\}} m_{r_j}. \quad (13)$$

We do not assume that $m_{\min} = m_{\max}$ therefore for a model with $m_{r_j} < m_{\max}$ the “redundant” entries in c_i and y_I are “ignored”.

To each ODE model we associate an observable function $O_{r_j} : \mathbb{R}^{m_{r_j}} \rightarrow \mathbb{R}^o$ which describes an experimental observation explained by that model where o denotes the dimension of the experimental observation. The expected observation $O_{r_j}^i$ of the r_j -th model at time point i is given by

$$O_{r_j}^i := O_{r_j}(y_{r_j}^i). \quad (14)$$

Let O^{t_i} denote an observation at measurement time point t_i . By assuming that the measurements at successive time points t_i are independent with normally distributed error vectors $\epsilon_{r_j}^i \in \mathbb{R}^m$ with zero mean and variance functions $v_{r_j}(O_{r_j}^i, t_i, \theta_{r_j})^2$ we get for the regression models

$$O^{t_i} = O_{r_j}^i + \epsilon_{r_j}^i \quad (15)$$

the model probability densities $f_{r_j}(\cdot; \cdot)$ at measurement time point t_i given by

$$f_{r_j}(O_{t_i}; O_{r_j}^i) = \frac{1}{\sqrt{2\pi}|v_{i,r_j}|} e^{-\frac{1}{2}(O_{r_j}^i - O^{t_i})^T V_{r_j}^i (O_{r_j}^i - O^{t_i})}, \quad (16)$$

with $|v_{i,r_j}| := \prod_{k=1}^o v_{r_j}^k(O_{r_j}^i, t_i, \theta_{r_j})$, where $v_{r_j}^k(O_{r_j}^i, t_i, \theta_{r_j})$ denotes the k -th entry of the square root of the variance functions $v_{r_j}(O_{r_j}^i, t_i, \theta_{r_j})^2$, and diagonal matrices $V_{r_j}^i \in \mathbb{R}^{o \times o}$ with diagonal entries $[V_{r_j}^i]_{kk} := (1/v_{r_j}^k(O_{r_j}^i, t_i, \theta_{r_j}))^2$.

We generally allow different error models. The error models might dependent on the expected observations $O_{r_j}^i$, the time t_i and possibly on parameters θ_{r_j} .

For the sake of notational simplicity we define

$$f_{r_j}(O^{t_i}) := f_{r_j}(O^{t_i}; O_{r_j}^i). \quad (17)$$

For a full measurement run containing n measurement time points we get the probability density models

$$f_{r_j}(O) := \prod_{i=1}^n f_{r_j}(O^{t_i}). \quad (18)$$

However, by assuming such a model probability distribution we still allow that two measurements are separated by a time span less than ΔT .

To overcome this problem we extend the probability spaces $\Omega_i = \mathbb{R}^o$ of a measurement at one measurement time point by one-element-containing sets \mathcal{N}_i to

$$\tilde{\Omega}_i = \Omega_i \cup \mathcal{N}_i \quad (19)$$

where $\tilde{\Omega}_i$ is the disjoint union of Ω_i and \mathcal{N}_i . The element of the set \mathcal{N}_i with measure $P(\mathcal{N}_i) \in [0, 1]$ represents the event “no measurement”, i.e. $\tilde{O}_{t_i} \in \mathcal{N}_i \Leftrightarrow$ “no measurement performed at time point t_i ”.

In order to derive measures on $\tilde{\Omega}_i$, $i = 1, \dots, n$ that allow for a density function representation according to the Radon-Nikodym theorem (Theorem 1), we introduce the Heaviside-function

$$\mathcal{H} : \mathbb{R}_+ \longrightarrow [0, 1] \quad (20)$$

with

$$\mathcal{H}(t_i) = \begin{cases} 1 & \text{if } t_i - t_{i-1} \geq \Delta T \\ 0 & \text{if } t_i - t_{i-1} < \Delta T \end{cases} \quad (21)$$

By use of this Heaviside-function and σ -algebras \mathcal{F}_i , where \mathcal{F}_i contains the Lebesgue measurable sets on Ω_i and additionally the union of these with the set \mathcal{N}_i , we define probability spaces $(\tilde{\Omega}_i, \mathcal{F}_i, \tilde{P}_{i,r_j})$ with measures

$$\tilde{P}_{i,r_j} : E_i \in \mathcal{F}_i \mapsto \tilde{P}_{i,r_j}(E_i) \in [0, 1]. \quad (22)$$

Three cases have to be distinguished: 1. $E_i \subset \Omega_i$, 2. $E_i \subset \mathcal{N}_i$, 3. $E_i \cap \Omega_i \neq \emptyset$ and $E_i \cap \mathcal{N}_i \neq \emptyset$.

For case one with $E_i \subset \Omega_i$ we set

$$\tilde{P}_{i,r_j}(E_i) := \mathcal{H}(t_i) \int_{E_i} f_{r_j}(O^{t_i}) dO^{t_i}. \quad (23)$$

For case two with $E_i \subset \mathcal{N}_i$ we set

$$\tilde{P}_{i,r_j}(E_i) := 1 - \mathcal{H}(t_i). \quad (24)$$

For case three with $E_i \cap \Omega_i \neq \emptyset$ and $E_i \cap \mathcal{N}_i \neq \emptyset$ we set

$$\tilde{P}_{i,r_j}(E_i) := \mathcal{H}(t_i) \int_{E_i \cap \Omega_i} f_{r_j}(O^{t_i}) dO^{t_i} + (1 - \mathcal{H}(t_i)). \quad (25)$$

By introducing these modifications the probability models based on measures \tilde{P}_{i,r_j} do not depend on measurements which are performed in less than ΔT time after the previous measurement any more.

To take into account that a species concentration perturbation can only be applied if no measurement is done at the same time, the same procedure is repeated with the additional Heaviside-function

$$\tilde{\mathcal{H}}(c_i) = \begin{cases} 0 & \text{if } c_i > 0 \\ 1 & \text{if } c_i = 0. \end{cases} \quad (26)$$

The measures \tilde{P}_{i,r_j} are defined in the same way as above replacing $\mathcal{H}(t_i)$ by $\mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i)$.

For specific r_1 and r_2 inserting the probability models \tilde{P}_{i,r_1} and respectively \tilde{P}_{i,r_2} into the KL-divergence (Definition 3) using $\lambda_i := \tilde{P}_{i,r_2}$ and the additivity of the KL divergence for independent events one gets the following expression

$$\begin{aligned} \mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) = \sum_{i=1}^n & \left[\int \mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i) f_{r_2}(O^{t_i}) \log \left\{ \frac{\mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i) f_{r_2}(O^{t_i})}{\mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i) f_{r_1}(O^{t_i})} \right\} dO^{t_i} + \right. \\ & \left. (1 - \mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i)) \cdot \log \frac{(1 - \mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i))}{(1 - \mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i))} \right], \end{aligned} \quad (27)$$

where $c_n := 0$. With $\log(1) = 0$ this simplifies to

$$\mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) = \sum_{i=1}^n \mathcal{H}(t_i)\tilde{\mathcal{H}}(c_i) \int f_{r_2}(O^{t_i}) \cdot \log \left\{ \frac{f_{r_2}(O^{t_i})}{f_{r_1}(O^{t_i})} \right\} dO^{t_i}. \quad (28)$$

By inserting the normal distribution (16) in (28) one gets

$$\begin{aligned} \mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) &= \sum_{i=1}^n \mathcal{H}(t_i) \tilde{\mathcal{H}}(c_i) \cdot \\ &\quad \int f_{r_2}(O^{t_i}) \cdot \log \left\{ \frac{\frac{1}{\sqrt{2\pi|v_{i,r_2}|}} e^{-\frac{1}{2}(O_{r_2}^i - O^{t_i})^T V_{r_2}^i (O_{r_2}^i - O^{t_i})}}{\frac{1}{\sqrt{2\pi|v_{i,r_1}|}} e^{-\frac{1}{2}(O_{r_1}^i - O^{t_i})^T V_{r_1}^i (O_{r_1}^i - O^{t_i})}} \right\} dO^{t_i}. \end{aligned} \quad (29)$$

This is equivalent to

$$\mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) = \sum_{i=1}^n \mathcal{H}(t_i) \tilde{\mathcal{H}}(c_i) \left(\sum_{k=1}^o \log \frac{v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1})}{v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2})} + A_k^i \right), \quad (30)$$

with

$$\begin{aligned} A_k^i &:= \frac{1}{2} \int \lfloor f_{r_2}(O^{t_i}) \rfloor_k \cdot \\ &\quad \left[-\frac{1}{(v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2} \left(\lfloor O_{r_2}^i \rfloor_k^2 - 2 \lfloor O_{r_2}^i \rfloor_k \lfloor O^{t_i} \rfloor_k + \lfloor O^{t_i} \rfloor_k^2 \right) + \right. \\ &\quad \left. \frac{1}{(v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1}))^2} \left(\lfloor O_{r_1}^i \rfloor_k^2 - 2 \lfloor O_{r_1}^i \rfloor_k \lfloor O^{t_i} \rfloor_k + \lfloor O^{t_i} \rfloor_k^2 \right) \right] \lfloor dO^{t_i} \rfloor_k. \end{aligned} \quad (31)$$

and where $\lfloor O \rfloor_k$ gives the k -th entry of the observation vector O . A_k^i reduces using the well known moments of the normal distribution to

$$\begin{aligned} A_k^i &= \frac{1}{2} \left[-\frac{1}{(v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2} \left(\lfloor O_{r_2}^i \rfloor_k^2 - 2 \lfloor O_{r_2}^i \rfloor_k^2 + \lfloor O_{r_2}^i \rfloor_k^2 + (v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2 \right) \right. \\ &\quad \left. + \frac{1}{(v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1}))^2} \left(\lfloor O_{r_1}^i \rfloor_k^2 - 2 \lfloor O_{r_1}^i \rfloor_k \lfloor O_{r_2}^i \rfloor_k + \lfloor O_{r_2}^i \rfloor_k^2 + (v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2 \right) \right]. \end{aligned} \quad (32)$$

This further simplifies to

$$A_k^i = \frac{1}{2} \left[-1 + \frac{\left(\lfloor O_{r_1}^i \rfloor_k^2 - 2 \lfloor O_{r_1}^i \rfloor_k \lfloor O_{r_2}^i \rfloor_k + \lfloor O_{r_2}^i \rfloor_k^2 + (v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2 \right)}{(v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1}))^2} \right]. \quad (33)$$

Substituting A_k^i back into (30) we get

$$\begin{aligned} \mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) &= \sum_{i=1}^n \mathcal{H}(t_i) \tilde{\mathcal{H}}(c_i) \sum_{k=1}^o \left(\log \left(\frac{v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1})}{v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2})} \right) + \right. \\ &\quad \left. \frac{1}{2} \left[-1 + \frac{\left(\lfloor O_{r_1}^i \rfloor_k - \lfloor O_{r_2}^i \rfloor_k \right)^2 + (v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2}{(v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1}))^2} \right] \right). \end{aligned} \quad (34)$$

This reduces to

$$\begin{aligned} \mathcal{I}(P_{r_2} : P_{r_1}, \mathcal{O}_1) = & \frac{1}{2} \sum_{i=1}^n \mathcal{H}(t_i) \tilde{\mathcal{H}}(c_i) \cdot \\ & \left(\sum_{k=1}^o \left[\frac{(v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2}))^2 + (\lfloor O_{r_2}^i \rfloor_k - \lfloor O_{r_1}^i \rfloor_k)^2}{(v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1}))^2} \right. \right. \\ & \left. \left. - 2 \log \left(\frac{v_{r_2}^k(O_{r_2}^i, t_i, \theta_{r_2})}{v_{r_1}^k(O_{r_1}^i, t_i, \theta_{r_1})} \right) \right] - o \right). \end{aligned} \quad (35)$$

This criterion has to be maximized with respect to the initial concentration vector y_I , the measurement time point vector t and the system perturbation vector c , thus $\xi := (y_I, t, c) \in \Xi \subset \mathbb{R}^d$.

For our optimal experimental design we generally start with a large number of measurement time points. By use of the Heaviside functions the number of measurement time points is reduced such that for $t_i - t_{i-1} < \Delta T$ the corresponding measurement time point is “turned off”.

These Heaviside-functions $\mathcal{H}(\cdot)$ and $\tilde{\mathcal{H}}(\cdot)$ can be replaced by any appropriate continuously differentiable switching functions with range space $[0, 1]$.

It should be noted that we assume that we have the same time discretization for measurements of different species and the addition of further species quantities. This assumption is practical especially for the application to in vitro experiments performed by biologists. For introducing arbitrary generic controls we need a more general formulation of time schemes, i.e. simultaneously time schemes which are independent of each other. One is associated with the controls, others may be associated with distinct observables which might be measured independently. The incorporation to the presented framework is subject of further work.

3 Solution of the max-min optimization problem

We formally state now the experimental design optimization problem \mathbf{P}_Θ :

$$\max_{(\tau, \xi) \in \Xi \subset \mathbb{R}^{d+1}} \tau \quad (36)$$

subject to

$$\begin{aligned} \min_{\substack{\theta_{r_1} \in \Theta_{r_1} \\ \theta_{r_2} \in \Theta_{r_2}}} \quad & \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi) - \tau \geq 0, \quad r_1 \in \{1, \dots, M_1\}, \quad r_2 \in \{1, \dots, M_2\}, \\ & \sum_{i=1}^n \Delta t_i = T_{\text{end}}, \\ & y_I^{\min} \leq y_I \leq y_I^{\max}, \\ & 0 \leq \Delta t \leq t^{\max}, \\ & 0 \leq c \leq c^{\max}, \end{aligned}$$

with $\xi := (y_I, \Delta t, c) \in \mathbb{R}^d$, $\Delta t_i := t_i - t_{i-1}$ and $\Theta := \{(\Theta_{r_1}, \Theta_{r_2}) | r_j \in \{1, \dots, M_j\}\}$. The auxiliary variable $\tau \in \mathbb{R}$ is used to transform the maxmin optimization problem (8) to a maximization problem with an infinite number of inequality constraints. The remaining constraints model the feasible range of experimental setups.

To solve optimization problem (36) numerically by applying efficient derivative based algorithms we replace the Heaviside functions $\mathcal{H}(t_i)$ and $\tilde{\mathcal{H}}(c_i)$ in (35) by continuously differentiable approximations, parametrized hyperbolic tangent functions of the form

$$\mathcal{H}'(t_i) = \frac{\tanh(\frac{6(\Delta t_i - b_1)}{a_1}) + 1}{2} \quad \text{and} \quad \tilde{\mathcal{H}}'(c_i) = \frac{\tanh(-\frac{6(c_i - b_2)}{a_2}) + 1}{2}. \quad (37)$$

The parameters $a_{1,2}$ characterize the width of the transition region between 0 and 1. The parameters $b_{1,2}$ determine the center of the transition region (see Figure 1). By setting the parameters in an adequate way arbitrarily close approximations of the Heaviside functions can be generated. A different approach to handle the discontinuous Heaviside functions would be to introduce binary variables and treat the resulting problem as Mixed Integer Nonlinear Programming problem. The drawback of this approach is that its solution can become very expensive. There seems to be little theoretical work in literature on Mixed Integer maxmin problems and an efficient solution strategy is not obvious in that case.

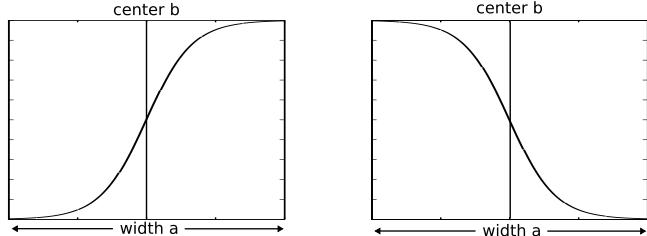


Figure 1: Switching functions: the left switching function is used to guarantee that only one measurement is done at a time point, the right one is used to guarantee that if a perturbation is done at a time point no measurement is done at the same time point.

In literature, problems as (36) fall into the class of semi-infinite inequality and equality constrained optimization problems (**SIECP**) [38].

Several methods to solve such **SIECP** problems are available, an overview can be found in [21, 38]. We choose the method of outer approximation [41, 42, 38], whose origin can be traced back to cutting plane methods for convex problems [38]. This approach is beneficial in the presence of a complex inner problem, in our case the robustification against the parameters θ_{r_1} and θ_{r_2} . The outer approximation algorithm iteratively solves discretized finite counterparts $\mathbf{P}_{\tilde{\Theta}^N}$ of the semi-infinite problem \mathbf{P}_Θ in each step N , successively refining the discretization $\tilde{\Theta}^N$ until a sufficient approximation of the original problem \mathbf{P}_Θ is reached. For problem (36) this means that in each iteration of the outer approximation

scheme Θ_{r_1} and Θ_{r_2} are replaced by finite approximations $\tilde{\Theta}_{r_1}^N$ and respectively $\tilde{\Theta}_{r_2}^N$ with $\tilde{\Theta}^N := \{(\tilde{\Theta}_{r_1}^N, \tilde{\Theta}_{r_2}^N) | r_j \in \{1, \dots, M_j\}\}$. This relation between the semi-infinite problem and an infinite sequence of finite problems can be formalized in the theory of consistent approximations and epi-convergence [35, 36, 37, 38].

We use a modified version of Algorithm 3.6.4 in [38] where “Step 1.”, the calculation of augmenting vectors $\tilde{\theta}_{r_1}^{N+1}$ and $\tilde{\theta}_{r_2}^{N+1}$ to construct

$$\tilde{\Theta}_{r_1}^{N+1} := \{\tilde{\theta}_{r_1}^{N+1}\} \cup \tilde{\Theta}_{r_1}^N \quad \text{and} \quad \tilde{\Theta}_{r_2}^{N+1} := \{\tilde{\theta}_{r_2}^{N+1}\} \cup \tilde{\Theta}_{r_2}^N, \quad (38)$$

is realized by

$$(\tilde{\theta}_{r_1}^{N+1}, \tilde{\theta}_{r_2}^{N+1}) := \arg \min_{\substack{\theta_{r_1} \in \Theta_{r_1} \\ \theta_{r_2} \in \Theta_{r_2}}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_N), \quad (39)$$

with ξ_N denoting a locally optimal design of the previous problem $\mathbf{P}_{\tilde{\Theta}^N}$. The algorithmic scheme is as follows:

Algorithm 1.

- *Data.* Choose $\xi_0 \in \Xi$ and a sequence $\{\epsilon_N\}_{N=1}^\infty$ with $\epsilon_N > 0$ and $\epsilon_N \downarrow 0$.
- *Step 0.* Set $N = 1$, set $\tilde{\Theta}_{r_j}^0 := \emptyset$.
- *Step 1.* Calculate $\tilde{\Theta}^N$ according to (38) and (39).
- *Step 2.* Calculate approximate solution of $\mathbf{P}_{\tilde{\Theta}^N}$ such that

$$\Psi_{\tilde{\Theta}_N}((\tau_N, \xi_N)) \geq -\epsilon_N, \quad (40)$$

and the equality and inequality constraints in problem $\mathbf{P}_{\tilde{\Theta}^N}$ are fulfilled up to ϵ_N .

- *Step 3.* Replace N by $N + 1$, and goto Step 1.

$\Psi_{\tilde{\Theta}_N}(\cdot) : \mathbb{R}^{d+1} \rightarrow \mathbb{R}_{\leq 0}$ denotes the optimality function associated to problem $\mathbf{P}_{\tilde{\Theta}^N}$, see Theorem 2.2.24 in [38]. The optimality function $\Psi_{\tilde{\Theta}_N}(\cdot)$ is always non positive and directly related to the first order generalized Karush-Kuhn-Tucker (KKT) conditions, i.e. $\Psi_{\tilde{\Theta}_N}((\tau, \xi)) = 0$ if evaluated at a generalized KKT point, see Theorem 2.2.19 in [38].

Assuming that $\mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi)$ and $\nabla_\xi \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi)$ are Lipschitz continuous on bounded sets with respect to ξ and θ_{r_j} and Θ_{r_j} are compact any accumulation point of Algorithm 1 fulfills the generalized KKT conditions, compare Theorem 3.6.5 in [38].

To calculate $\tilde{\theta}_{r_j}^{N+1}$ in Step 1. of Algorithm 1 we use on heuristic base a simple random search approach coupled to a local optimization method, i.e. we have randomly generated P different start values in Θ_{r_j} , from which we have started the local optimization method for parameter estimation. The best value out

of the P trials is chosen to augment the set $\tilde{\Theta}_{r_j}^N$. Of course there are more sophisticated approaches to search for a global minimum for a review see e.g. [4], but at this point an effective calculation of Step 1. of Algorithm 1 was not our primary goal. For the local parameter optimization we use the same optimization method as for Step 2. in Algorithm 1.

In our implementation we use a fixed ϵ at the desired final accuracy to solve problem $\mathbf{P}_{\tilde{\Theta}^N}$ in every loop of Step 2. of Algorithm 1, i.e. $\epsilon_N = \epsilon$, $N > 0$. In that way Step 1. of Algorithm 1 gives a worst case estimate of the KL divergence $\mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_N)$ for the current design ξ_N with respect to θ_{r_j} . Therefore for practical application the algorithm can be stopped if the worst case estimate of KL divergence for the current design ξ_N is big enough although no local optima might be achieved during optimization. As stopping criterion of Algorithm 1 we use:

Algorithmic Stop Criterion.

Stop after Step 1. of Algorithm 1, if

$$\begin{aligned} \delta \geq \min_{\substack{r_1 \in \{1, \dots, M_1\} \\ r_2 \in \{1, \dots, M_2\}}} \min_{\substack{\theta_{r_1} \in \tilde{\Theta}_{r_1}^{N-1} \\ \theta_{r_2} \in \tilde{\Theta}_{r_2}^{N-1}}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_{N-1}) - \\ \min_{\substack{r_1 \in \{1, \dots, M_1\} \\ r_2 \in \{1, \dots, M_2\}}} \min_{\substack{\theta_{r_1} \in \Theta_{r_1} \\ \theta_{r_2} \in \Theta_{r_2}}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_{N-1}) =: \Delta_{RG}, \end{aligned} \quad (41)$$

where δ is a small positive constant, then consider $(\tilde{\theta}_{r_1}^{N+1}, \tilde{\theta}_{r_2}^{N+1})$, $r_1 \in \{1, \dots, M_1\}$, $r_2 \in \{1, \dots, M_2\}$ and ξ_N as (approximate) solutions of problem \mathbf{P}_{Θ} , else goto Step 2. and calculate a new design ξ_{N+1} .

This stop criterion has also been used in [42, 39]. We call the distance Δ_{RG} given by (41), robustification gap.

3.1 Numerical solution of the problem $\mathbf{P}_{\tilde{\Theta}^N}$

We have implemented the resulting optimization problem in a multiple shooting setup (see for example [45, 14, 13]). In our multiple shooting setup the whole integration interval $[0, T_{\text{end}}]$ is subdivided into several subintervals by introducing auxiliary multiple shooting node variables $s_{r_j, i, l}$, $j \in \{1, 2\}$, $r_j \in \{1, \dots, M_j\}$, $i \in \{1, \dots, n\}$, $l \in \{1, \dots, N\}$, on each of which an independent initial value problem is solved. Each end point of a subinterval corresponds to one measurement time point. Matching conditions which enter the optimization problem as additional equality constraints assure continuity of the state trajectory from one subinterval to the next.

To incorporate the perturbations c matching conditions

$$s_{r_j, i, l} - y_{r_j}(t_{i-1}, t_i, s_{r_j, i-1, l}, \tilde{\theta}_{r_j}^l) = 0, \quad (42)$$

$s_{r_j, i, l}$, denoting the multiple shooting nodes with $s_{r_j, 0, l} = y_I$ are modified to

$$\begin{aligned} s_{r_j, i, l} - y_{r_j}(t_{i-1}, t_i, s_{j, i-1, l}, \tilde{\theta}_{r_j}^l) = c_i, \quad i \in \{1, \dots, n-1\}, \\ s_{r_j, n, l} - y_{r_j}(t_{n-1}, t_n, s_{j, n-1, l}, \tilde{\theta}_{r_j}^l) = 0. \end{aligned} \quad (43)$$

A graphical scheme of the multiple shooting setup is shown in Figure 2. Instead of evaluating the objective functional (35) by use of the values $y_{r_j}^i$, given by the solution of the initial value problem (11), depending on the parameters $\tilde{\theta}_{r_j}$, (35) is evaluated by use of the auxiliary multiple shooting node variables $s_{r_j,i,l}$ replacing the values $y_{r_j}^i$ with $s_{r_j,i,l}$ respectively. The dependency of (35) on $s_{r_j,i,l}$ is indicated by $\mathcal{I}(s_{r_1,\cdot,l}, s_{r_2,\cdot,l})$.

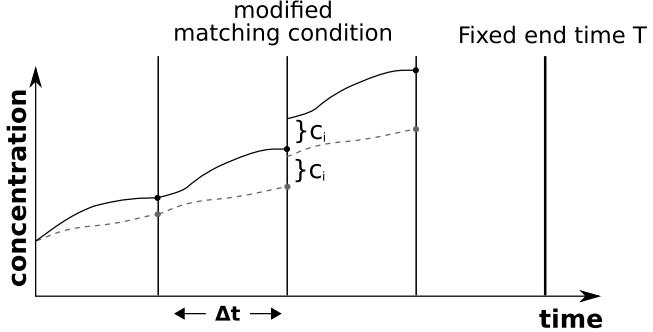


Figure 2: Scheme of the multiple shooting setup for computing the experimental design to discriminate two models. A dot denotes one measurement time point. The black solid line corresponds to model 1 and the gray dashed one to model 2.

The overall optimization problem can be stated as

$$\max_{\tau, y_I, \Delta t, c, s} \tau \quad (44)$$

subject to

$$\begin{aligned} \frac{dy_{r_j}}{dt} &= f_{r_j}^{\text{rhs}}(y, \tilde{\theta}_{r_j}^l), \quad t \in [t_{i-1}, t_i], \quad y_{r_j}(t_{i-1}) := s_{r_j,i-1,l} \\ s_{r_j,i,l} - y_{r_j}(t_{i-1}, t_i, s_{r_j,i-1,l}, \tilde{\theta}_{r_j}^l) &= c_i, \\ s_{r_j,n,l} - y_{r_j}(t_{n-1}, t_n, s_{r_j,n-1,l}, \tilde{\theta}_{r_j}^l) &= 0, \\ s_{r_j,0,l} &= y_I, \\ y_I^{\min} \leq y_I \leq y_I^{\max}, & \\ 0 \leq \Delta t \leq t^{\max}, & \\ 0 \leq c \leq c^{\max}, & \\ s_{r_j,i,l}^{\min} \leq s_{r_j,i,l} \leq s_{r_j,i,l}^{\max}, & \\ \sum_{i=1}^n \Delta t_i &= T_{\text{end}}, \\ \mathcal{I}(s_{r_1,\cdot,l}, s_{r_2,\cdot,l}) - \tau &\geq 0, \quad r_1 \in \{1, \dots, M_1\}, \quad r_2 \in \{1, \dots, M_2\}, \end{aligned} \quad (45)$$

with $j \in \{1, 2\}$, $r_j \in \{1, \dots, M_j\}$, $i \in \{1, \dots, n\}$, $l \in \{1, \dots, N\}$.

We have implemented this problem within the interior point optimization package IPOPT [50, 51], using the linear solver MA27 [23]. Usually for the solution of the KKT system within the direct multiple shooting ansatz the linear system

is solved utilizing tailored structure-exploitation, e.g., condensing or high rank updates. See for example [30]. Since speed aspects are not our primary concern we rely on the sparse solver MA27 instead of developing a tailored solver for this problem class at the current stage.

All derivatives up to second order, which are used for the calculations of the Hessian needed for a robust performance of IPOPT, are calculated by automatic differentiation using CppAD, [9, 8].

For the solution of the differential equations within the optimization problem, which are commonly stiff in chemical and biochemical applications, we have implemented a fully variable step, variable order (order 1 to 6), Backward differentiation formulae (BDF) method, based on Nordsiek array polynomial interpolation similar to the EPISODE BDF method by Byrne and Hindmarsh [16], but with the step size selection strategy of Calvo and Rández [17].

For the generation of sensitivities we have adopted the sophisticated principles of internal numerical differentiation developed by Albersmeyer and Bock [2, 1] in forward and adjoint mode.

The idea of this principle is instead of calculating the sensitivities by use of the sensitivity differential equation, to directly differentiate the BDF integration scheme by automatic differentiation, which we implemented using CppAD [9, 8].

According to some notes in the PhD thesis of Albersmeyer [1] we also have implemented the possibility to control the step size scheme not only by the local truncation error of the nominal trajectory but as well by the local truncation error of the sensitivities generated by the forward mode of automatic differentiation with respect to the sensitivity differential equation, which has shown by numerical experience to improve the robustness of the optimization approach. A different approach would be to use collocation, i.e. to incorporate a full discretization of the ODEs into the optimization problem, see e.g. [11]. Since the kinetic ODE systems in the focus of our applications are usually stiff, we prefer adaptive time integration.

3.2 Stabilizing homotopy method for subsequent $\mathbf{P}_{\tilde{\Theta}^{N+1}}$

By solving the subsequent optimization problems $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ with an interior point code like IPOPT [50, 51] initialized with primal and dual variables of the previous problem or with primal variables only, one often observes that the new solution may differ significantly from the previous. This is due to the fact that the solution of the previous problem $\mathbf{P}_{\tilde{\Theta}^N}$ is infeasible for $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ and thus the algorithm tries to find a feasible state before it proceeds to find a new optimum. This behavior is not desired in the context of an outer approximation algorithm, because convergence of the algorithm may be slowed down significantly. This circumstance originates from a jumping between vicinities of distinct local maxima of problem (36). The discretization $\tilde{\Theta}^N$ of the robustification space may not be equally adequate for different local maxima. To overcome this problem we have implemented a heuristic homotopy method to gradually introduce the additional constraints

$$g_{r_1, r_2}(\tau, \xi)_{N+1} := \mathcal{I}(s_{r_1, \cdot, N+1}, s_{r_2, \cdot, N+1}) - \tau \geq 0, \quad (46)$$

of problem $\mathbf{P}_{\tilde{\Theta}^{N+1}}$. We replace $g_{r_1, r_2}(\tau, \xi)_{N+1}$ by

$$\tilde{g}_{r_1, r_2}(\tau, \xi; \kappa)_{N+1} := \mathcal{I}(s_{r_1, \cdot, N+1}, s_{r_2, \cdot, N+1}) - \tau + (1 - \kappa)\rho \geq 0, \quad (47)$$

with homotopy parameter $\kappa \in [0, 1]$ and ρ is a constant which has to be set such that $\tilde{g}_{r_1, r_2}(\tau, \xi; \kappa)_{N+1}$ are inactive for $\kappa = 0$ at the initial design ξ_N . We choose ρ to be

$$\rho := K \cdot \max_{\substack{r_1 \in \{1, \dots, M_1\} \\ r_2 \in \{1, \dots, M_2\}}} \left(\min_{\substack{\theta_{r_1} \in \tilde{\Theta}_{r_1}^N \\ \theta_{r_2} \in \tilde{\Theta}_{r_2}^N}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_N) - \min_{\substack{\theta_{r_1} \in \Theta_{r_1} \\ \theta_{r_2} \in \Theta_{r_2}}} \mathcal{I}(P_{r_2}(\theta_{r_2}) : P_{r_1}(\theta_{r_1}), \mathcal{O}_1; \xi_N) \right). \quad (48)$$

K is a save guard factor we set empirically to $K = 1.4$, which worked well in practice for our examples. For $\kappa = 0$ the augmented optimization problem should be easily solvable within a few iterations by performing a warm start from the solution of the previous problem. By increasing the homotopy parameter to $\kappa = 1$ the additional constraint is gradually introduced, which leads to a sequence of easily solvable subproblems whose solutions stay in the vicinity of the solution of the previous problem $\mathbf{P}_{\tilde{\Theta}^N}$. A similar homotopy strategy can be found e.g. in [40].

4 Numerical results

We have applied the algorithm developed in Section 2 and 3 to two example problems for which we present results in the following section. For the purpose of illustration we restrict ourself to the case that each hypothesis comprise only one model whereby we assume that only the alternative hypothesis is composite. We also assume that each species is “directly” measurable. We treat model 1 as null hypothesis and model 2 as alternative hypothesis.

4.1 Discriminating design for two models describing glycolytic oscillations

In the first test case for model discrimination we implemented the following models for glycolytic oscillations as described in [20].

Model 1 is an allosteric enzyme model with positive feedback under cooperativity and linear product sink. The differential equations for model 1 are given by

$$\begin{aligned} \frac{d\alpha_1}{dt} &= \nu - \sigma\phi(\alpha_1, \gamma_1), \\ \frac{d\gamma_1}{dt} &= q_1\sigma\phi(\alpha_1, \gamma_1) - k_s\gamma_1, \\ \phi(\alpha_1, \gamma_1) &= \frac{\alpha_1(1 + \alpha_1)(1 + \gamma_1)^2}{L_1 + (1 + \alpha_1)^2(1 + \gamma_1)^2}. \end{aligned}$$

Model 2 is an allosteric model with positive feedback in the absence of cooperativity and the product sink is represented by Michaelis-Menten kinetics. The differential equations for Model 2 are given by

$$\begin{aligned}\frac{d\alpha_2}{dt} &= \nu - \phi(\alpha_2, \gamma_2), \\ \frac{d\gamma_2}{dt} &= q_2\phi(\alpha_2, \gamma_2) - \frac{r_s\gamma_2}{\mu + \gamma_2}, \\ \phi(\alpha_2, \gamma_2) &= \frac{\alpha_2(1 + \gamma_2)}{L_2 + (1 + \alpha_2)(1 + \gamma_2)}.\end{aligned}$$

$\alpha_{1,2}$ denotes the species concentration of the substrate $\gamma_{1,2}$ that of the product. For both models the inflow parameter ν is the same and fixed to the value $\nu = 0.22$. It represents the inflow of substrate to the experimental system, a CSTR (continuously stirred tank reactor).

The parameters σ , q_1 , k_s and L_1 of model 1 are regarded as known. Their values are given in Table 1, the parameters q_2 , r_s , μ and L_2 of model 2 are regarded as unknown and subject to robustification. For the permitted parameter range see Table 1.

Model 1				Model 2			
σ	q_1	k_s	L_1	q_2	r_s	μ	L_2
0.92	2.01	0.11	17206.10	$[10^{-7}, 100]$	$[10^{-7}, 100]$	$[10^{-7}, 100]$	$[100, 300]$

Table 1: Parameter values for the glycolytic oscillation models.

For simplicity we consider the homoscedastic case with equal variances, i.e. $v_1 = v_2 = \sigma^2$. In this case $\mathcal{I}(P_1 : P_2, \mathcal{O}_1)$ reduces to,

$$\mathcal{I}(P_1 : P_2, \mathcal{O}_1) = \sum_{i=1}^n \mathcal{H}'(t_i) \widetilde{\mathcal{H}'}(c_i) ((\alpha_1^i - \alpha_2^i)^2 + (\gamma_1^i - \gamma_2^i)^2). \quad (49)$$

For this test case the homotopy strategy as presented in Section 3.2 is only applied if the robustification gap $\Delta_{RG} < 0.1$, then the successive problem $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ is calculated by use of the homotopy strategy with 30 homotopy steps, i.e. $\kappa_h = h/30$, $h \in \{1, \dots, 30\}$. Otherwise the problem $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ is solved without homotopy strategy. For each subsequent problem $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ the solution of problem $\mathbf{P}_{\tilde{\Theta}^N}$ is used as initial guess.

We first present a robust design without the possibility to disturb the system by adding species at later time points.

The design is calculated within a fixed time window i.e. $T_{\text{end}} = 400$. 100 equally spaced possible measurement points are defined in the initial state of the optimization procedure, the distance vector Δt between the time points is subject to design and each entry is restricted to $\Delta t_i \in [10^{-7}, 10^{19}]$, $i \in \{1, \dots, 100\}$. The disturbance vectors c_i are set to $c_i = 0$, $i \in \{1, \dots, 99\}$ and are fixed to model

the fact that no species disturbance is allowed.

The initial species concentrations which are also subject to experimental design are restricted to $\alpha_I \in [10^{-7}, 25]$ and $\gamma_I \in [10^{-7}, 25]$. The initial values were set to $\alpha_I = 15$ and $\gamma_I = 2$. The parameters of the switching functions $\mathcal{H}'(t_i)$ are chosen as $a_1 = 20.0$ and $b_1 = 10.0$. The parameters of the switching functions $\widetilde{\mathcal{H}}'(c_i)$ are chosen as $a_2 = 0.05$ and $b_2 = 0.025$. The algorithmic settings are summarized in Table 2.

Optimization settings			Integrator settings	
P	δ	IPOPT-tol: Step 1./Step 2. $10^{-10}/10^{-8}$	relTol/absTol $10^{-12}/10^{-12}$	relTolSens/absTolSens $10^{-12}/10^{-12}$

Table 2: On the left hand side the optimization settings are listed comprising the IPOPT stopping tolerances for Step 1. and Step 2. of Algorithm 1 and on the right hand side the integration tolerances for the nominal trajectory and the first order sensitivities. We use the IPOPT option “honor_original_bounds=no” for Step 1. and Step 2. of Algorithm 1.

A plot of the functions α_1, α_2 and γ_1, γ_2 in the initial state and for the solution of problem $\mathbf{P}_{\tilde{\Theta}^1}$ are shown in Figure 3. A plot for the same functions with the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}^1}$ after the next robustification step is shown in Figure 4. The final design is also shown in Figure 4. A plot of the

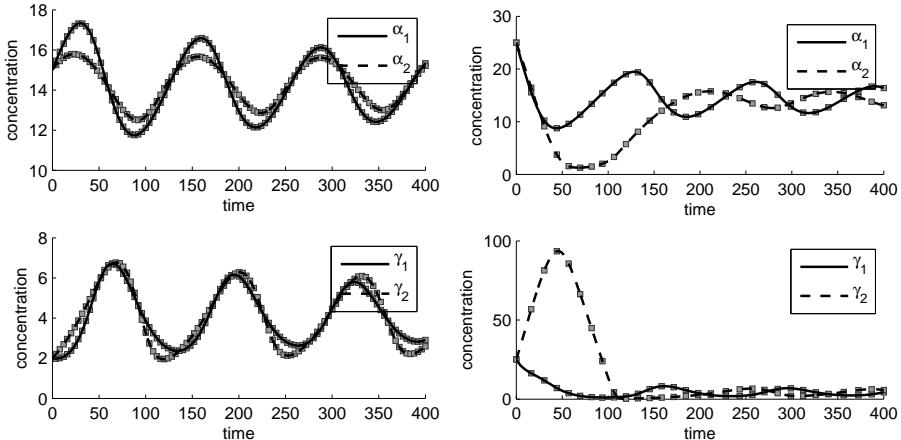


Figure 3: The model functions α_1, α_2 and γ_1, γ_2 are shown before the optimization procedure (left) and after the optimization procedure of problem $\mathbf{P}_{\tilde{\Theta}^1}$ (right) for the glycolytic design setup without the possibility to disturb the system. One square represents one measurement time point.

robustification gap Δ_{RG} and as well for the objective value of problem $\mathbf{P}_{\tilde{\Theta}^N}$ for each iteration N of Algorithm 1 are shown in Figure 5. A selection of design variables as solutions of problem $\mathbf{P}_{\tilde{\Theta}^N}$ is shown in Figure 6(left).

In a second scenario we additionally allow for species perturbations. In this

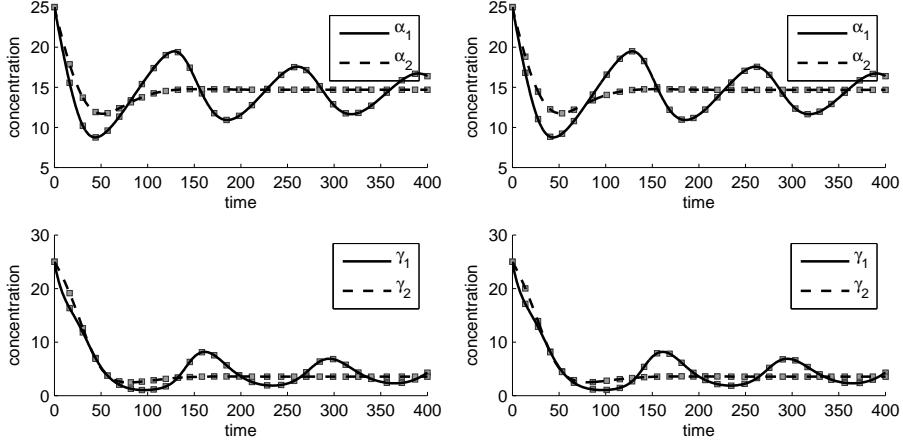


Figure 4: The model functions α_1, α_2 and γ_1, γ_2 are shown for the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}^1}$ after the next robustification step (left) and for the final design (right) for the glycolytic design setup without the possibility to disturb the system. One square represents one measurement time point.

new scenario at the 21th, 41th, 61th and 81th measurement time points, the system can get disturbed by additional species quantities. The free vectors c_i , $i \in \{21, 41, 61, 81\}$ are constrained by $c_i \in [10^{-7}, 10]$. The initial values are set to $c_i = 1$. The remaining conditions are as before, however we change the time vector bound constraints for $i \in \{1, 6, 11, 21, 26, 31, 41, 46, 51, 61, 66, 71, 81\}$ to $\Delta t_i \in [8, 10^{19}]$ and the initial state to $\Delta t_i = 15$. The bounds for the remaining entries are as before, and the remaining measurement time points were equally spaced.

A plot of the functions α_1, α_2 and γ_1, γ_2 in the initial state and for the solution of problem $\mathbf{P}_{\tilde{\Theta}^1}$ are shown in Figure 7. A plot for the same functions with the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}^1}$ after the next robustification step is shown in Figure 8. The final design is also shown in Figure 8. A plot of the robustification gap Δ_{RG} and the objective value of problem $\mathbf{P}_{\tilde{\Theta}^N}$ for each iteration N of Algorithm 1 are shown in Figure 9. A selection of design variables as solutions of problem $\mathbf{P}_{\tilde{\Theta}^N}$ is shown in Figure 6(right).

4.2 Discriminating design for two models describing signal sensing in *dictyostelium discoideum*

The second test case is the discrimination of two models describing the chemo-tactic response in the amoeba *Dictyostelium discoideum* as presented in [33] using the framework presented in Section 2 and 3. The two models describe the adaption mechanism observed when amoebae encounter the chemoattractant cAMP [31], see figure 10. For both models, a chemotaxis response regulator R gets activated (R^*) by an activator enzyme A , when a cAMP ligand S appears. But the deactivating mechanism determined by the interaction with an inhibitor molecule I differs for both models. Both models comprise mass action kinetics in form of ODE.

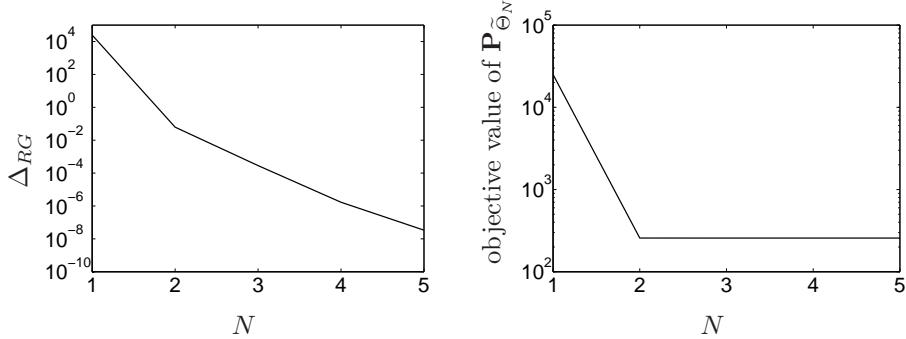


Figure 5: In the left figure the robustification gap Δ_{RG} is plotted versus the number of iterations N of Algorithm 1 and in the right figure the objective value of problem $\mathbf{P}_{\tilde{\Theta}_N}$ is shown for the glycolytic design setup without the possibility to disturb the system.

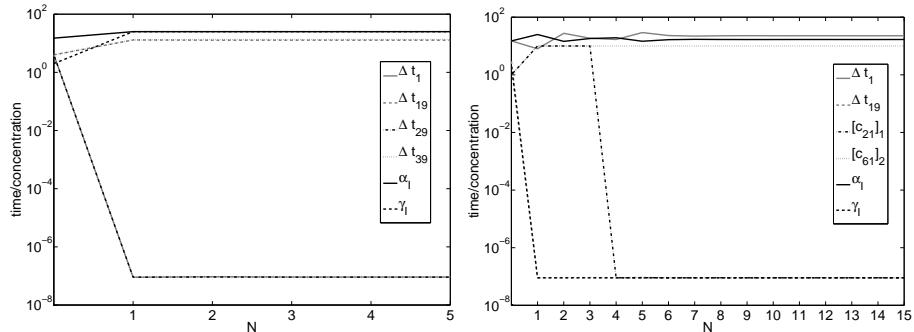


Figure 6: A selection of design variables as solutions of problem $\mathbf{P}_{\tilde{\Theta}_N}$ for the glycolytic design setup without the possibility to disturb the system (left) and with the possibility to disturb the system (right) are shown.

In model 1 the activator enzyme as well as the inhibitor enzyme are regulated by the external signal, which is proportional to the cAMP concentration S . The overall model in this case is given by,

$$\begin{aligned} \dot{A}_1 &= -k_{-a}A_1 + k_aS_1 \\ \dot{I}_1 &= -k_{-i}I_1 + k_{i_1}S_1 \\ \dot{R}_1^* &= -(k_rA_1 + k_{-r}I_1)R_1^* + k_rR_T A_1, \end{aligned} \tag{50}$$

where k_{-a} , k_a , k_{-i} , k_{i_1} , k_r and k_{-r} are the mass action rate constants and $R_T := R^* + R$ is the total amount of the response regulator.

In model 2 the inhibitory molecule I is activated through the indirect action of activator A instead of direct activation by sensing ligand binding. The overall

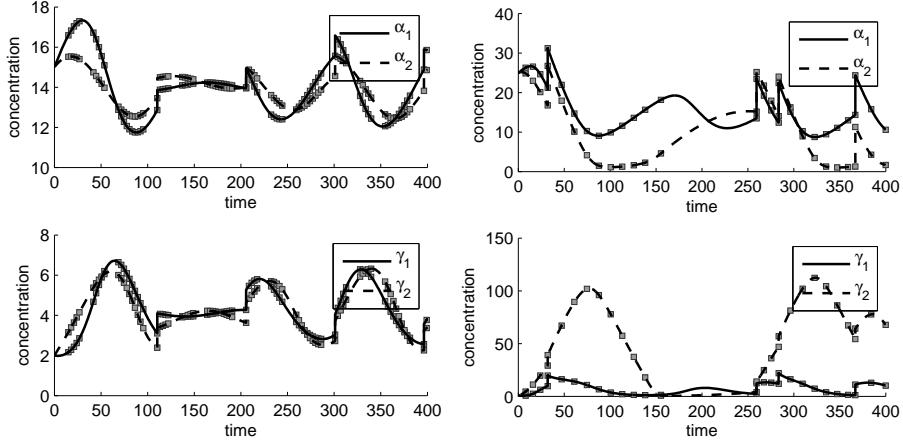


Figure 7: The model functions α_1, α_2 and γ_1, γ_2 are shown before the optimization procedure (left) and after the optimization procedure of problem $\mathbf{P}_{\tilde{\Theta}^1}$ (right) for the glycolytic design setup with the possibility to disturb the system. One square represents one measurement time point.

model in this case is given by,

$$\begin{aligned}\dot{A}_2 &= -k_{-a}A_2 + k_aS_2 \\ \dot{I}_2 &= -k_{-i}I_2 + k_{i_2}A_2 \\ \dot{R}_2^* &= -(k_rA_2 + k_{-r}I_2)R_2^* + k_rR_T A_2,\end{aligned}\tag{51}$$

where k_{-a} , k_a , k_{-i} , k_{i_2} , k_r and k_{-r} are the mass action rate constants and $R_T := R^* + R$ is the total amount of the response regulator.

For modeling details we refer to [33]. We have extended these systems of ordinary differential equations by an additional state corresponding to the cAMP ligand S with $\dot{S} = 0$. By allowing species concentration perturbations c only to the state S we can mimic a piecewise constant control of the system by the cAMP ligand S .

The experimental design parameters are the initial species concentrations of the four states namely, A_I , I_I , R_I , S_I , the measurement time points t and the species concentration perturbation c with respect to S . We discard the condition that either a measurement or a perturbation can be performed since in that setting by use of the perturbations c we mimic a piecewise constant input control S and therefore that restriction seems unnatural. Again for simplicity we consider the homoscedastic case with equal variances i.e. $v_1 = v_2 = \sigma^2$, where $\mathcal{I}(P_1 : P_2, \mathcal{O}_1)$ reduces now to

$$\mathcal{I}(P_1 : P_2, \mathcal{O}_1) = \sum_{i=1}^n \mathcal{H}'(t_i) \left((A_1^i - A_2^i)^2 + (I_1^i - I_2^i)^2 + (R_1^{*,i} - R_2^{*,i})^2 \right). \tag{52}$$

The parameters k_{-a} , k_a , k_{-i} , k_{i_1} , k_r , k_{-r} and R_T are regarded as known and fixed, their values are given in Table 3. Parameter k_{i_2} is regarded as un-

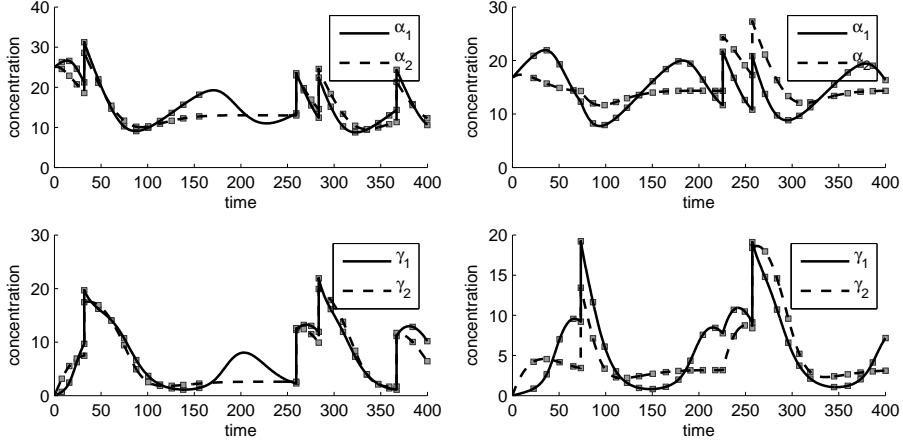


Figure 8: The model functions α_1, α_2 and γ_1, γ_2 are shown for the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}^1}$ after the next robustification step (left) and for the final design (right) for the glycolytic design setup with the possibility to disturb the system. One square represents one measurement time point.

k_{-a}	k_a	k_{-i}	k_{i_1}	k_r	k_{-r}	R_T
2.0	3.0	0.1	1.0	1.0	1.0	23/30

Table 3: Parameter values for the fix values within model 1 and model 2.

known and subject to robustification. The range of the parameter k_{i_2} is set to $k_{i_2} \in [0, 2]$.

The optimal design is calculated within a fixed time window with $T_{\text{end}} = 100$. 100 equally spaced possible measurement points are defined in the initial state of the optimization procedure. The distance vector Δt between time points is subject to design and each entry is restricted to $\Delta t_i \in [10^{-7}, 10^{19}], i \in \{1, \dots, 100\}$. The free perturbation vectors $c_i, i \in \{11, 21, 31, 41, 51, 61, 71, 81, 91\}$ are not restricted. The initial values are set to $c_i = 0, i \in \{11, 21\}, c_{31} = 0.3, c_i = -0.48, i \in \{41, 61, 81\}$ and $c_i = 0.48, i \in \{51, 71, 91\}$.

The initial species concentrations which are also subject to the experimental design are restricted to $S_I \in [0.01, 0.5]$, $A_I \in [10^{-7}, 1]$, $I_I \in [10^{-7}, 1]$ and $R_I \in [10^{-7}, 1]$. The initial values are set to $S_I = 0.2$, $A_I = 1.0$, $I = 10^{-4}$ and $R = 10^{-4}$. The multiple shooting intermediate variables for the species S are restricted to $s_i \in [0.01, 0.5]$ to restrict the piecewise constant control to this interval. The parameters of the switching functions $\mathcal{H}'(t_i)$ are chosen as $a_1 = 5.0$ and $b_1 = 2.5$. The algorithmic settings are summarized in Table 4.

With these design conditions we start the optimization procedure twice. First by use of the homotopy strategy for successive problems $\mathbf{P}_{\tilde{\Theta}^{N+1}}$ with 10 homotopy steps.

Since the “discriminating power” of the experimental setup is very low in this case, i.e. the deviation between the two models is small, we plot the distance functions $(S_1 - S_2)$, $(A_1 - A_2)$, $(I_1 - I_2)$ and $(R_1 - R_2)$ for the initial state and

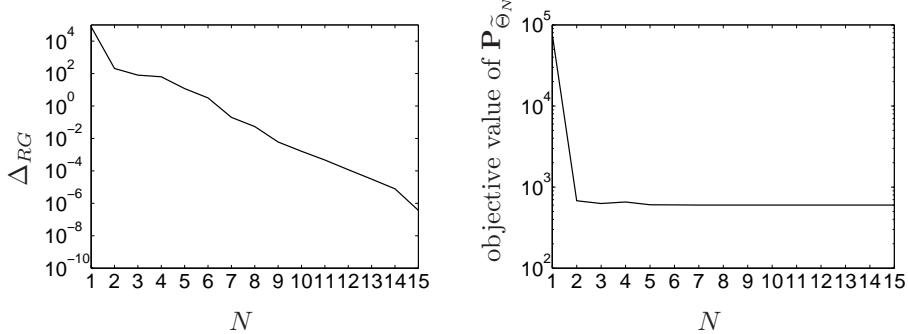


Figure 9: In the left figure the robustification gap Δ_{RG} is plotted versus the number of iterations N of Algorithm 1 and in the right figure the objective value of problem $\mathbf{P}_{\tilde{\Theta}_N}$ is shown for the glycolytic design setup with the possibility to disturb the system.

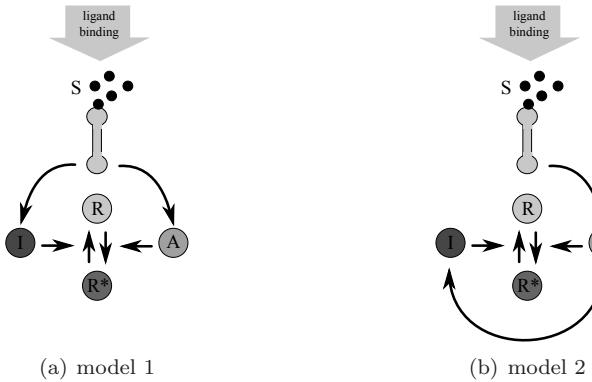


Figure 10: Two models of the signal system of the Dictyostelium amoeba.

for the solution of problem $\mathbf{P}_{\tilde{\Theta}_1}$ in Figure 11. A plot for the same functions with the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}_1}$ after the next robustification step is shown in Figure 12. The final design is also shown in Figure 12.

A plot of the robustification gap Δ_{RG} for each iteration N of Algorithm 1 is shown in Figure 13 (left). A plot of the objective value of problem $\mathbf{P}_{\tilde{\Theta}_N}$ for each iteration N of Algorithm 1 is shown in Figure 14 (left). A selection of design variables as solutions of problem $\mathbf{P}_{\tilde{\Theta}_N}$ is shown in Figure 15 (left).

Second we calculate the design without the homotopy strategy. We experience huge jumps in the final objective value of problem $\mathbf{P}_{\tilde{\Theta}_N}$ for subsequent iterations N of Algorithm 1. This is due to the fact that the final design of the former problem $\mathbf{P}_{\tilde{\Theta}_N}$ is an infeasible starting point for the successive problem $\mathbf{P}_{\tilde{\Theta}_{N+1}}$ in the interior point solution strategy. First the optimizer tries to force the iterates back into the feasible region and afterwards the new central path leads to a different design.

For this case a plot of the robustification gap Δ_{RG} for each iteration N of Algorithm 1 is shown in Figure 13 (right). A plot of the objective value of problem

Optimization settings			Integrator settings	
P	δ	IPOPT-tol: Step 1./Step 2. $10^{-10}/10^{-11}$	relTol/absTol $10^{-14}/10^{-14}$	relTolSens/absTolSens $10^{-14}/10^{-14}$

Table 4: On the left hand side the optimization settings are listed comprising the IPOPT stopping tolerances for Step 1. and Step 2. of Algorithm 1 and on the right hand side the integration tolerances for the nominal trajectory and the first order sensitivities. We use the IPOPT option “honor_original_bounds=no” for Step 1. and Step 2. of Algorithm 1.

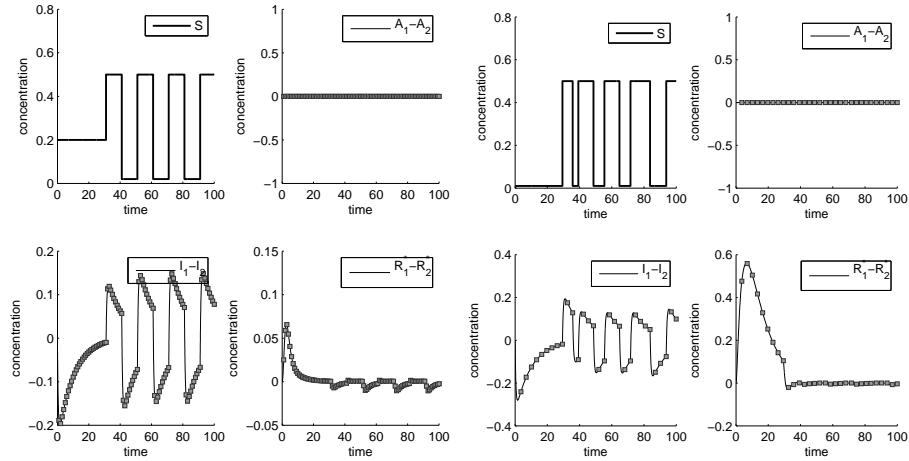


Figure 11: The model variable distance functions $(S_1 - S_2)$, $(A_1 - A_2)$, $(I_1 - I_2)$ and $(R_1 - R_2)$ are shown before the optimization procedure (left) and after the optimization procedure of problem $\mathbf{P}_{\tilde{\Theta}_1}$ (right) for two models describing signal sensing in *dictyostelium discoideum*. One square represents one measurement time point.

$\mathbf{P}_{\tilde{\Theta}_N}$ for each iteration N of Algorithm 1 is shown in Figure 14 (right). A selection of design variables as solutions of problem $\mathbf{P}_{\tilde{\Theta}_N}$ is shown in Figure 15 (right). As one can clearly see, the homotopy strategy helps to considerably stabilize Algorithm 1.

5 Conclusion

We present a framework for the robust computation of optimal experimental designs for the purpose of model discrimination. The theoretical framework as well as the numerical realization by utilization of an outer approximation algorithm are worked out. A strategy for the numerical stabilization of the algorithm by use of a homotopy approach is suggested. The optimization procedure is successfully exemplified on two biological model systems. In our examples we clearly found that the homotopy approach is significantly superior to a cold start of successive design problems $\mathbf{P}_{\tilde{\Theta}_{N+1}}$. For the first test case, the discrimi-

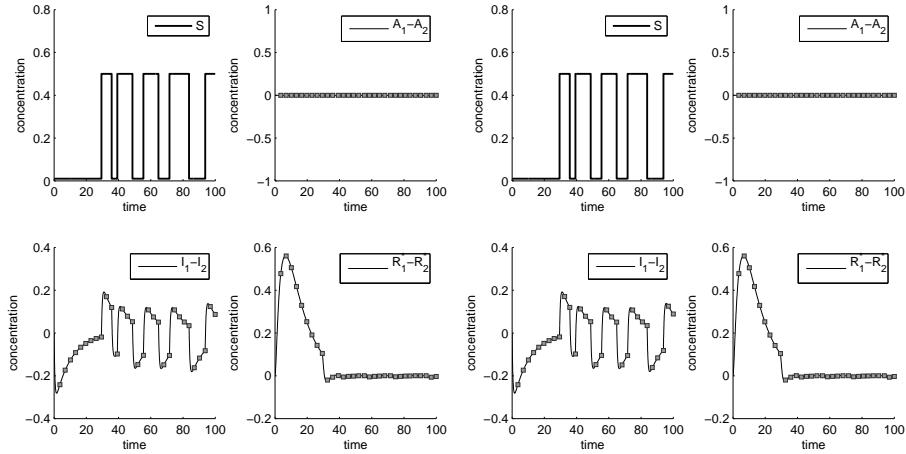


Figure 12: The model variable distance functions $(S_1 - S_2)$, $(A_1 - A_2)$, $(I_1 - I_2)$ and $(R_1 - R_2)$ are shown for the same solution design as for problem $\mathbf{P}_{\tilde{\Theta}^1}$ after the next robustification step and for the final design (right) for two models describing signal sensing in *dictyostelium discoideum*. One square represents one measurement time point.

nation of two models describing glycolytic oscillations, the outer approximation scheme completely fails to reach the desired accuracy δ without homotopy strategy. For the second test case, the discrimination of two models describing signal sensing in *dictyostelium discoideum*, the outer approximation scheme also fails without warmstart, however the homotopy strategy also works with only two homotopy steps (not presented in this paper). This indicates the need of a step size strategy for reasons of efficiency which will be a next step in our work.

6 Acknowledgement

The authors thank the anonymous reviewers for helpful comments and suggestions.

The authors gratefully acknowledge the Freiburg Initiative for Systems Biology (FRISYS), part of the BMBF FORSYS systems biology initiative, the Freiburg excellence cluster Centre for Biological Signalling Studies (BIOS), the Helmholtz alliance Systems Biology of Cancer and the Nephage initiative (BMBF Gerontosys II) for various support and funding.

References

[1] Jan Albersmeyer. *Adjoint-based algorithms and numerical methods for sensitivity generation and optimization of large scale dynamic systems*. PhD thesis, University of Heidelberg, Heidelberg, December 2010.

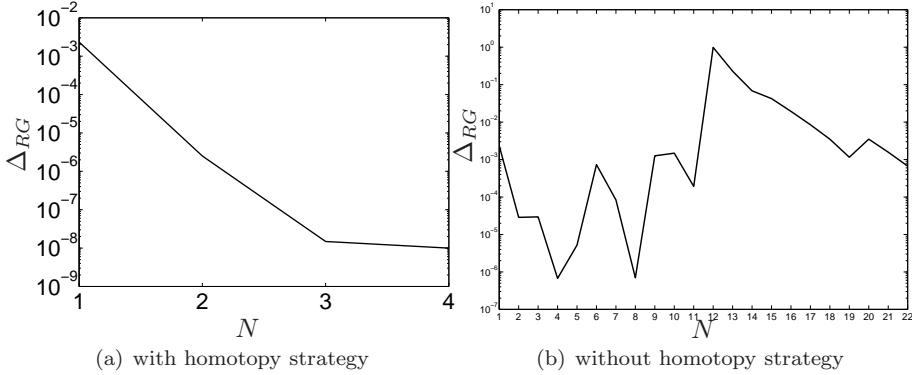


Figure 13: The robustification gap Δ_{RG} is plotted versus the number of iterations N of Algorithm 1 for the setup with two models describing signal sensing in *dictyostelium discoideum*, with homotopy strategy (left) and without homotopy strategy (right).

- [2] Jan Albersmeyer and Hans Georg Bock. Sensitivity generation in an adaptive BDF-method. In *Modeling, Simulation and Optimization of Complex Processes: Proceedings of the Third International Conference on High Performance Scientific Computing*. Springer, 2008.
- [3] Joshua F Apgar, Jared E Toettcher, Drew Endy, Forest M White, and Bruce Tidor. Stimulus design for model selection and validation in cell signaling. *PLoS Computational Biology*, 4(2):e30, 02 2008.
- [4] J. S. Arora, O. A. Elwakeil, A. I. Chahande, and C. C. Hsieh. Global optimization methods for engineering applications: A review. *Structural and Multidisciplinary Optimization*, 9:137–159, 1995. 10.1007/BF01743964.
- [5] A. C. Atkinson and V. V. Fedorov. The design of experiments for discriminating between two rival models. *Biometrika*, 62(1):57–70, 1975.
- [6] E. Balsa-Canto, A. A. Alonso, and J. R. Banga. Computational procedures for optimal experimental design in biological systems. *IET Systems Biology*, 2(4):163–172, July 2008.
- [7] I. Bauer, H. G. Bock, S. Körkel, and J. P. Schlöder. Numerical methods for optimum experimental design in DAE systems. *Journal of Computational and Applied Mathematics*, 120:1–25, 2000.
- [8] Bradley M. Bell. Automatic differentiation software cppad., 2010.
- [9] Bradley M. Bell and James V. Burke. Algorithmic differentiation of implicit functions and optimal values. In Christian H. Bischof, H. Martin Bücker, Paul D. Hovland, Uwe Naumann, and J. Utke, editors, *Advances in Automatic Differentiation*, pages 67–77. Springer, Berlin, 2008.
- [10] Joseph P. Bernacki and Regina M. Murphy. Model discrimination and mechanistic interpretation of kinetic data in protein aggregation studies. *Biophysical Journal*, 96:2871–2887, 2009.

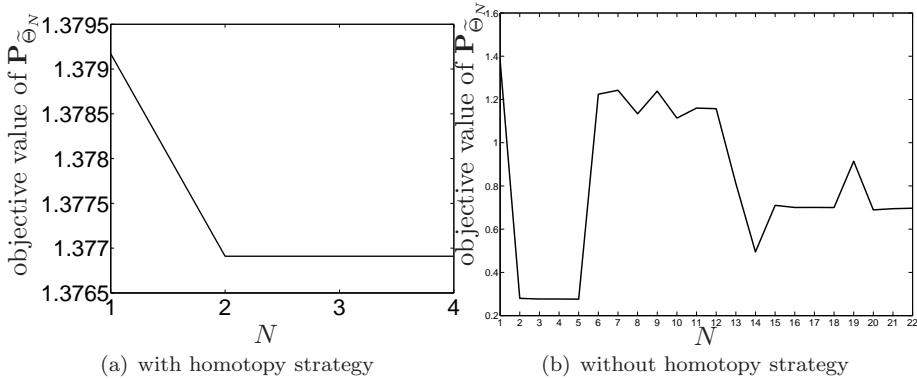


Figure 14: The objective value of problem $\mathbf{P}_{\tilde{\Theta}^N}$ is plotted versus the number of iterations N of Algorithm 1 for the setup with two models describing signal sensing in *dictyostelium discoideum*, with homotopy strategy (left) and without homotopy strategy (right).

- [11] Lorenz T. Biegler, Arturo M. Cervantes, and Andreas Wächter. Advances in simultaneous strategies for dynamic process optimization. *Optimization, Chemical Engineering Science*, 57:575–593, 2001.
- [12] Patrick Billingsley. *Probability and Measure*. John Wiley & Sons Inc, 1986.
- [13] Hans Georg Bock. Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. In *Bonner Mathematische Schriften*, volume 183. University of Bonn, 1987.
- [14] Hans Georg Bock and Karl J. Plitt. A multiple shooting algorithm for direct solution of optimal control problems. In *Proceedings of the Ninth IFAC World Congress, Budapest*. Pergamon, Oxford, 1984.
- [15] Kenneth P. Burnham and David R. Anderson. *Model Selection and Multimodel inference: A practical information-theoretic approach*. Springer, 2002.
- [16] G. D. Byrne and A. C. Hindmarsh. A polyalgorithm for the numerical solution of ordinary differential equations. *ACM Transactions on Mathematical Software*, 1(1):71–96, 1975.
- [17] M. Calvo, J. I. Montijano, and L. Rández. On the change of step size in multistep codes. *Numerical Algorithms*, 4:283–304, 1993.
- [18] Herman Chernoff. Large-sample theory: Parametric case. *The Annals of Mathematical Statistics*, 27(1):pp. 1–22, 1956.
- [19] M. J. Cooney and K. A. McDonald. Optimal dynamic experiments for bioreactor model discrimination. *Applied Microbiology and Biotechnology*, 43:826–837, 1995.
- [20] Albert Goldbeter. *Biochemical oscillations and cellular rhythms: The molecular bases of periodic and chaotic behaviour*. Cambridge University Press, 1996.

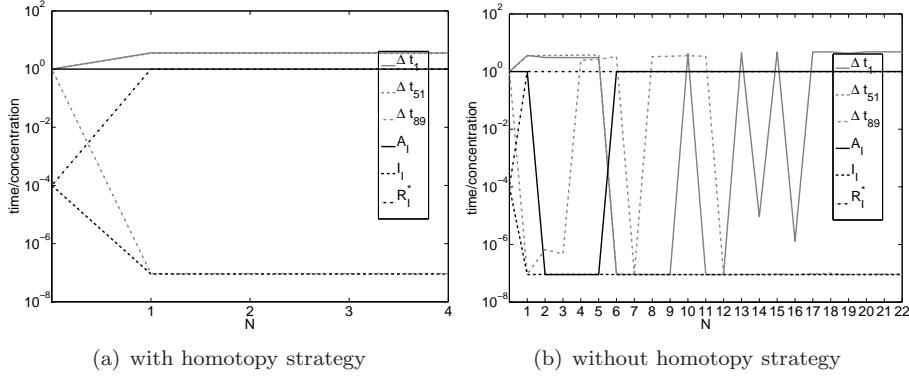


Figure 15: A selection of design variables calculated as solution of problem $\mathbf{P}_{\bar{\mathcal{Q}}^N}$ is plotted versus the number of iterations N of Algorithm 1 for the setup with two models describing signal sensing in *dictyostelium discoideum*, on the left with homotopy strategy and on the right without homotopy strategy.

- [21] R. Hettich and K. O. Kortanek. Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35(3):pp. 380–429, 1993.
- [22] R. Horn. Statistical methods for model discrimination. applications to gating kinetics and permeation of the acetylcholine receptor channel. *Bioophysical Journal*, 51:255–263, 1987.
- [23] HSL. A collection of fortran codes for large-scale scientific computation. See <http://www.hsl.rl.ac.uk>, 2007.
- [24] Rishi Jain, Andrea L. Knorr, Joseph Bernacki, and Ranjan Srivastava. Investigation of bacteriophage ms2 viral dynamics using model discrimination analysis and the implications for phage therapy. *Biotechnology Progress*, 22(6):1650–1658, 2006.
- [25] S. Körkel, I. Bauer, H. G. Bock, and J. P. Schlöder. A sequential approach for nonlinear optimum experimental design in DAE systems. In F. Keil, W. Mackens, H. Voss, , and J. Werther, editors, *Scientific Computing in Chemical Engineering II*, volume 2. Springer Verlag, Berlin, 1999.
- [26] A. Kremling, S. Fischer, K. Gadkar, F. J. Doyle, T. Sauter, E. Bullinger, F. Allgöwer, and E. D. Gilles. A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Research*, 14(9):1773–1785, September 2004.
- [27] Clemens Kreutz and Jens Timmer. Systems biology: experimental design. *FEBS Journal*, 276:923–942, 2009.
- [28] Solomon Kullback. *Information Theory and Statistics*. Dover Publications Inc., 1997.
- [29] Laurence Lacey and Adrian Dunne. The design of pharmacokinetic experiments for model discrimination. *Journal of Pharmacokinetics and Pharmacodynamics*, 12:351–365, 1984.

[30] Daniel B. Leineweber. *Efficient Reduced SQP Methods for the Optimization of Chemical Processes Described by Large Sparse DAE Models*. PhD thesis, University of Heidelberg, 1998.

[31] A Levchenko and PA Iglesias. Models of eukaryotic gradient sensing: Application to chemotaxis of amoebae and neutrophils. *Biophysical Journal*, 82:50–63, 2002.

[32] J. López-Fidalgo, C. Tommasi, and P. C. Trandafir. An optimal experimental design criterion for discriminating between non-normal models. *Journal of the Royal Statistical Society Series B*, 69(2):231–242, 2007.

[33] Bence Melykuti, Elias August, Antonis Papachristodoulou, and Hana El-Samad. Discriminating between rival biochemical network models: three approaches to optimal experiment design. *BMC Systems Biology*, 4(1):38, 2010.

[34] Jay I. Myung and Mark A. Pitt. Optimal experimental design for model discrimination. *Psychological review*, 116(3):499–518, July 2009.

[35] E. Polak. On the convergence of optimization algorithms. *Rev. Française Informat. Recherche Opérationnelle*, 3(16):17–34, 1969.

[36] E. Polak. On the mathematical foundations of nondifferentiable optimization in engineering design. *SIAM Review*, 29(1):pp. 21–89, 1987.

[37] E. Polak. On the use of consistent approximations in the solution of semi-infinite optimization and optimal control problems. *Mathematical Programming*, 62:385–414, 1993. 10.1007/BF01585175.

[38] Elijah Polak. *Optimization: Algorithms and Consistent Approximations*. Springer, 1997.

[39] Luc Pronzato and Eric Walter. Robust experiment design via maximin optimization. *Mathematical Biosciences*, 89(2):161 – 176, 1988.

[40] Victor Pérez, John Renaud, and Layne Watson. Homotopy curve tracking in approximate interior point optimization. *Optimization and Engineering*, 10:91–108, 2009. 10.1007/s11081-008-9042-6.

[41] D. Salmon. Minimax controller design. *Automatic Control, IEEE Transactions on*, 13(4):369 – 376, aug. 1968.

[42] Kiyotaka Shimizu and Eitaro Aiyoshi. Necessary conditions for min-max problems and algorithms by a relaxation procedure. *IEEE Transactions on Automatic Control*, 25(1):62–66, 1980.

[43] Dominik Skanda and Dirk Lebiedz. An optimal experimental design approach to model discrimination in dynamic biochemical systems. *Bioinformatics*, 26(7):939–945, 2010.

[44] W. E. Stewart, Y. Shon, and G. E. P. Box. Discrimination and goodness of fit of multiresponse mechanistic models. *AIChE Journal*, 44(6):1404–1412, 1998.

- [45] Josef Stoer and Roland Bulirsch. *Introduction to Numerical Analysis*. Number 12 in Texts in Applied Mathematics. Springer, New York, third edition, 2002.
- [46] C. Stricker, S. Redman, and D. Daley. Statistical analysis of synaptic transmission: model discrimination and confidence limits. *Biophysical Journal Of The Royal Statistical Society Series B*, 67:532–547, 1994.
- [47] R. Takors, W. Wiechert, and D. Weuster-Botz. Experimental design for the identification of macrokinetic models and model discrimination. *Biotechnol Bioeng*, 56(5):564–576, Dec 1997.
- [48] Jens Timmer, T. G. Müller, I. Swameye, O. Sandra, and U. Klingmüller. Modeling the nonlinear dynamics of cellular signal transduction. *International Journal of Bifurcation and Chaos*, 14(6):2069–2079, 2004.
- [49] D. Uciński and B. Bogacka. T-optimum designs for multiresponse dynamic heteroscedastic models. In A. Di Bucchianico and H. Lauter, editors, *Proc. of the 7th International Workshop on Model-Oriented Design and Analysis*, pages 191–199. Physica Verlag, 2004.
- [50] Andreas Wächter. *An Interior Point Algorithm for Large-Scale Nonlinear Optimization with Applications in Process Engineering*. PhD thesis, Carnegie Mellon University, 2002.
- [51] Andreas Wächter and Lorenz T. Biegler. On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1):25–57, 2006.